|  |
| --- |
| Close-up image showing the leaf-sides of two oversized books side-by-side on a bookshelf, with additional books in soft focus background |
| **MySQL Capstone project**  **E-Commerce Customer Churn Analysis** |
| |  |  |  | | --- | --- | --- | | P. Zahir hussain | [Date] | Data Science & Machine Learning | |

**Project Manager**

**Entri Elevate**

**Coding Mentor**

**Janani Lakshmi**

**Narayanan**

**Instructors**

**Archana S**

**Coordinator**

**Rafsal A**

**Introduction**

In the fast-paced world of e-commerce, retaining customers is as vital as acquiring new ones. Customer churn, the phenomenon where customers discontinue their business with a company, poses a significant challenge to e-commerce businesses, impacting profitability, brand loyalty, and market competitiveness. Understanding the intricacies of customer churn is crucial for organizations aiming to foster long-term relationships with their clientele.

This project, titled "E-Commerce Customer Churn Analysis," seeks to explore the patterns and factors associated with customer attrition in the e-commerce sector. By analyzing historical transactional data, the project aims to unveil the underlying drivers of churn and discern customer behaviors that lead to increased or decreased loyalty. Utilizing a comprehensive dataset encompassing various customer attributes—such as tenure, preferred payment methods, satisfaction scores, and purchase behaviors—our analysis intends to identify actionable insights to help e-commerce enterprises develop targeted retention strategies.

The objectives of this project are multifaceted. Firstly, we will conduct data cleaning to address missing values, inconsistencies, and outliers, ensuring the dataset's integrity. Subsequently, we will engage in data exploration and analysis to uncover critical trends and relationships, including the demographic distribution of churned customers, the impact of complaints on churn rates, and spending patterns among active and churned customers.

Furthermore, this analysis will involve creating new columns to enhance our understanding of customer status and engagement. By evaluating how different customer segments—such as demographics, payment preferences, and complaint behaviors—interact with churn, we hope to provide guidance for e-commerce providers to implement effective retention initiatives. Ultimately, our aim is to empower these businesses with the knowledge necessary to mitigate churn effectively, thereby enhancing customer satisfaction and ensuring sustained growth in an increasingly competitive landscape.

Through this comprehensive analysis, we aspire to bridge the gap between raw data and strategic decision-making, equipping e-commerce businesses with the tools they need to thrive in a dynamic environment.

**Area Of Study**

**Project Title: E-Commerce Customer Churn Analysis**

Problem Statement:

In the realm of e-commerce, businesses face the challenge of understanding customer

churn patterns to ensure customer satisfaction and sustained profitability. This project

aims to delve into the dynamics of customer churn within an e-commerce domain,

utilizing historical transactional data to uncover underlying patterns and drivers of

churn. By analyzing customer attributes such as tenure, preferred payment modes,

satisfaction scores, and purchase behavior, the project seeks to investigate and

understand the dynamics of customer attrition and their propensity to churn. The

ultimate objective is to equip e-commerce enterprises with actionable insights to

implement targeted retention strategies and mitigate churn, thereby fostering long-term

customer relationships and ensuring business viability in a competitive landscape.

Dataset Download:

https://drive.google.com/uc?export=download&id=1iKKCze\_Fpk2n\_g3BIZBiSjcDFdFcEn3D

Project Steps and Objectives:

Data Cleaning:

Handling Missing Values and Outliers:

➢ Impute mean for the following columns, and round off to the nearest integer if

required: WarehouseToHome, HourSpendOnApp, OrderAmountHikeFromlastYear,

DaySinceLastOrder.

➢ Impute mode for the following columns: Tenure, CouponUsed, OrderCount.

➢ Handle outliers in the 'WarehouseToHome' column by deleting rows where the

values are greater than 100.

Dealing with Inconsistencies:

➢ Replace occurrences of “Phone” in the 'PreferredLoginDevice' column and

“Mobile” in the 'PreferedOrderCat' column with “Mobile Phone” to ensure

uniformity.

➢ Standardize payment mode values: Replace "COD" with "Cash on Delivery" and

"CC" with "Credit Card" in the PreferredPaymentMode column.

Data Transformation:

Column Renaming:

➢ Rename the column "PreferedOrderCat" to "PreferredOrderCat".

➢ Rename the column "HourSpendOnApp" to "HoursSpentOnApp".

Creating New Columns:

➢ Create a new column named ‘ComplaintReceived’ with values "Yes" if the

corresponding value in the ‘Complain’ is 1, and "No" otherwise.

➢ Create a new column named 'ChurnStatus'. Set its value to “Churned” if the

corresponding value in the 'Churn' column is 1, else assign “Active”.

Column Dropping:

➢ Drop the columns "Churn" and "Complain" from the table.

Data Exploration and Analysis:

1. Retrieve the count of churned and active customers from the dataset.

2. Display the average tenure of customers who churned.

3. Calculate the total cashback amount earned by customers who churned.

4. Determine the percentage of churned customers who complained.

5. Find the gender distribution of customers who complained.

6. Identify the city tier with the highest number of churned customers whose

preferred order category is Laptop & Accessory.

7. Identify the most preferred payment mode among active customers.

8. List the preferred login device(s) among customers who took more than 10 days

since their last order.

9. List the number of active customers who spent more than 3 hours on the app.

10. Find the average cashback amount received by customers who spent at least 2

hours on the app.

11. Display the maximum hours spent on the app by customers in each preferred

order category.

12. Find the average order amount hike from last year for customers in each marital

status category.

13. Calculate the total order amount hike from last year for customers who are single

and prefer mobile phones for ordering.

14. Find the average number of devices registered among customers who used UPI as

their preferred payment mode.

15. Determine the city tier with the highest number of customers.

16. Find the marital status of customers with the highest number of addresses.

17. Identify the gender that utilized the highest number of coupons.

18. List the average satisfaction score in each of the preferred order categories.

19. Calculate the total order count for customers who prefer using credit cards and

have the maximum satisfaction score.

20. How many customers are there who spent only one hour on the app and days

since their last order was more than 5?

21. What is the average satisfaction score of customers who have complained?

22. How many customers are there in each preferred order category?

23. What is the average cashback amount received by married customers?

24. What is the average number of devices registered by customers who are not

using Mobile Phone as their preferred login device?

25. List the preferred order category among customers who used more than 5

coupons.

26. List the top 3 preferred order categories with the highest average cashback

amount.

27. Find the preferred payment modes of customers whose average tenure is 10

months and have placed more than 500 orders.

28. Categorize customers based on their distance from the warehouse to home such

as 'Very Close Distance' for distances <=5km, 'Close Distance' for <=10km,

'Moderate Distance' for <=15km, and 'Far Distance' for >15km. Then, display the

churn status breakdown for each distance category.

29. List the customer’s order details who are married, live in City Tier-1, and their

order counts are more than the average number of orders placed by all

customers.

30. a) Create a ‘customer\_returns’ table in the ‘ecomm’ database and insert the

following data:

ReturnID CustomerID ReturnDate RefundAmount

1001 50022 2023-01-01 2130

1002 50316 2023-01-23 2000

1003 51099 2023-02-14 2290

1004 52321 2023-03-08 2510

1005 52928 2023-03-20 3000

1006 53749 2023-04-17 1740

1007 54206 2023-04-21 3250

1008 54838 2023-04-30 1990

b) Display the return details along with the customer details of those who have

churned and have made complaints.

**Steps and Objectives**

Sure! Below is a description of the fields that are typically found in a dataset for e-commerce customer churn analysis. This can vary based on the specific dataset you're using, but here is a general outline of common fields and their descriptions:

**Field Descriptions**

| **Field Name** | **Description** |
| --- | --- |
| CustomerID | Unique identifier for each customer in the dataset. |
| Tenure | Duration (in months) the customer has been with the company. |
| PreferredLoginDevice | Device the customer prefers to log in with (e.g., Mobile Phone, Desktop). |
| PreferredPaymentMode | Customer's preferred method of payment for transactions (e.g., Cash on Delivery, Credit Card). |
| PreferredOrderCat | Categories of products that the customer prefers to order (e.g., Electronics, Clothing, Home Goods). |
| OrderCount | Total number of orders placed by the customer in the specified timeframe. |
| HourSpendOnApp | Total hours spent by the customer on the e-commerce application. |
| WarehouseToHome | Distance from the warehouse to the customer's home (often in kilometers). |
| DaySinceLastOrder | The number of days since the customer last placed an order. |
| OrderAmountHikeFromLastYear | Percentage increase or decrease in the order amount compared to the previous year. |
| SatisfactionScore | A score that indicates the customer's satisfaction level with their experiences (typically on a fixed scale, e.g., 1-10). |
| Churn | Binary indicator of churn status; 1 if the customer has churned, 0 if active. |
| Complain | Binary indicator of whether the customer has lodged a complaint; 1 if yes, 0 if no. |
| CashbackAmount | Total amount of cashback received by the customer. |
| CityTier | Classification of the customer's city based on economic characteristics (e.g., Tier 1, Tier 2, Tier 3). |
| Gender | Gender of the customer (e.g., Male, Female, Other). |
| MaritalStatus | Marital status of the customer (e.g., Single, Married). |
| CouponUsed | The number of coupons used by the customer, if applicable. |
| AddressesCount | The total number of addresses registered by the customer in their profile. |
| DevicesRegistered | The number of devices (e.g., smartphones, tablets) registered to the customer's account. |
| ComplaintReceived | A derived field indicating whether a complaint was received from the customer ("Yes" or "No"). |

**Note:**

The actual field names may vary, and there may be additional fields specific to your dataset. It's important to refer to the dataset documentation for precise definitions and additional context. The fields listed above are intended to give a comprehensive overview of what might be included in an e-commerce customer churn analysis dataset.

**Data Analysis SQL Query’s**

-- Project Title: E-Commerce Customer Churn Analysis

USE ecomm;

select \* from customer\_churn;

-- Data Cleaning:

-- Handling Missing Values and Outliers

-- Impute mean for the following columns, and round off to the nearest integer if

-- required: WarehouseToHome, HourSpendOnApp, OrderAmountHikeFromlastYear,

-- DaySinceLastOrder.

-- Disable safe update mode

SET SQL\_SAFE\_UPDATES = 0;

-- 1 WarehouseToHome

UPDATE customer\_churn SET WarehouseToHome = ROUND(

(SELECT AVG(WarehouseToHome)

FROM (SELECT WarehouseToHome FROM customer\_churn) AS temp

WHERE WarehouseToHome IS NOT NULL)

),

-- 2 HourSpendOnApp

HourSpendOnApp = ROUND(

(SELECT AVG(HourSpendOnApp)

FROM (SELECT HourSpendOnApp FROM customer\_churn) AS temp

WHERE HourSpendOnApp IS NOT NULL)

),

-- 3 OrderAmountHikeFromlastYear

OrderAmountHikeFromlastYear = ROUND(

(SELECT AVG(OrderAmountHikeFromlastYear)

FROM (SELECT OrderAmountHikeFromlastYear FROM customer\_churn) AS temp

WHERE OrderAmountHikeFromlastYear IS NOT NULL)

),

-- 4 DaySinceLastOrder

DaySinceLastOrder = ROUND(

(SELECT AVG(DaySinceLastOrder)

FROM (SELECT DaySinceLastOrder FROM customer\_churn) AS temp

WHERE DaySinceLastOrder IS NOT NULL)

);

select \* from customer\_churn;

-- ➢ Impute mode for the following columns: Tenure, CouponUsed, OrderCount.

-- 1 Impute mode for Tenure column

UPDATE customer\_churn SET Tenure = (

SELECT Tenure FROM (

SELECT Tenure, COUNT(\*) as count

FROM customer\_churn

WHERE Tenure IS NOT NULL

GROUP BY Tenure

ORDER BY count DESC

LIMIT 1) AS subquery

);

select \* from customer\_churn;

-- 2 Impute mode for CouponUsed column

UPDATE customer\_churn SET CouponUsed = (

SELECT CouponUsed FROM (

SELECT CouponUsed, COUNT(\*) as count

FROM customer\_churn

WHERE CouponUsed IS NOT NULL

GROUP BY CouponUsed

ORDER BY count DESC

LIMIT 1) AS subquery

);

select \* from customer\_churn;

-- 3 Impute mode for OrderCount column

UPDATE customer\_churn SET OrderCount = (

SELECT OrderCount FROM (

SELECT OrderCount, COUNT(\*) as count

FROM customer\_churn

WHERE OrderCount IS NOT NULL

GROUP BY OrderCount

ORDER BY count DESC

LIMIT 1) AS subquery

);

select \* from customer\_churn;

-- Handle outliers in the 'WarehouseToHome' column by deleting rows where the

-- values are greater than 100.

DELETE FROM customer\_churn

WHERE WarehouseToHome > 100;

select \* from customer\_churn;

-- Dealing with Inconsistencies:

-- ➢ Replace occurrences of “Phone” in the 'PreferredLoginDevice' column and

-- “Mobile” in the 'PreferedOrderCat' column with “Mobile Phone” to ensure uniformity.

UPDATE customer\_churn

SET PreferredLoginDevice = REPLACE(PreferredLoginDevice, 'Phone', 'Mobile')

WHERE PreferredLoginDevice LIKE '%Phone%';

select \* from customer\_churn;

UPDATE customer\_churn

SET PreferedOrderCat = REPLACE(PreferedOrderCat, 'Mobile', 'Mobile Phone')

WHERE PreferedOrderCat LIKE '%Mobile%';

select \* from customer\_churn;

-- Standardize payment mode values: Replace "COD" with "Cash on Delivery" and

-- "CC" with "Credit Card" in the PreferredPaymentMode column.

UPDATE customer\_churn SET PreferredPaymentMode = CASE

WHEN PreferredPaymentMode = 'COD' THEN 'Cash on Delivery'

WHEN PreferredPaymentMode = 'CC' THEN 'Credit Card'

ELSE PreferredPaymentMode

END

WHERE PreferredPaymentMode IN ('COD', 'CC');

select \* from customer\_churn;

-- Data Transformation:

-- Column Renaming:

-- Rename the column "HourSpendOnApp" to "HoursSpentOnApp"

ALTER TABLE customer\_churn

RENAME COLUMN HourSpendOnApp TO HoursSpentOnApp;

-- Creating new columns:

-- Create a new column named ‘ComplaintReceived’ with values "Yes" if the

-- corresponding value in the ‘Complain’ is 1, and "No" otherwise.

-- 1 Add the new column

ALTER TABLE customer\_churn

ADD COLUMN ComplaintReceived VARCHAR(30);

select \* from customer\_churn;

-- 2 Update the new column with values based on the Complain column

UPDATE customer\_churn SET ComplaintReceived = CASE

WHEN Complain = 1 THEN 'Yes'

ELSE 'No'

END;

select \* from customer\_churn;

-- 1 Add the new column

ALTER TABLE customer\_churn

ADD COLUMN ChurnStatus VARCHAR(10);

-- 2 Update the new column with values based on the Churn column

UPDATE customer\_churn

SET ChurnStatus = CASE

WHEN Churn = 1 THEN 'Churned'

ELSE 'Active'

END;

select \* from customer\_churn;

-- Column Dropping:

-- Drop the columns "Churn" and "Complain" from the table

ALTER TABLE customer\_churn

DROP COLUMN Churn,

DROP COLUMN Complain;

-- enable safe update mode

SET SQL\_SAFE\_UPDATES = 1;

-- Data Exploration and Analysis:

-- 1. Retrieve the count of churned and active customers from the dataset.

SELECT ChurnStatus, COUNT(\*) AS CustomerCount

FROM customer\_churn

GROUP BY ChurnStatus;

-- 2. Display the average tenure of customers who churned.

SELECT AVG(Tenure) AS AverageTenure

FROM customer\_churn

WHERE ChurnStatus = 'Churned';

-- 3. Calculate the total cashback amount earned by customers who churned.

SELECT SUM(CashbackAmount) AS TotalCashback

FROM customer\_churn

WHERE ChurnStatus = 'Churned';

-- 4. Determine the percentage of churned customers who complained.

SELECT (COUNT(CASE WHEN ComplaintReceived = 1 THEN 1 END) \* 100.0 / COUNT(\*)) AS PercentageComplained

FROM customer\_churn

WHERE ChurnStatus = 'Churned';

-- 5. Find the gender distribution of customers who complained.

SELECT Gender,COUNT(\*) AS ComplaintReceived

FROM customer\_churn

WHERE ComplaintReceived = 1

GROUP BY Gender;

use ecomm;

select \* from customer\_churn;

-- 6. Identify the city tier with the highest number of churned customers whose

-- preferred order category is Laptop & Accessory.

SELECT CityTier, COUNT(\*) AS ChurnedCustomerCount

FROM customer\_churn

WHERE ChurnStatus = 'Churned'

AND PreferedOrderCat = 'Laptop & Accessory'

GROUP BY CityTier

ORDER BY ChurnedCustomerCount DESC

LIMIT 2;

-- 7. Identify the most preferred payment mode among active customers.

SELECT PreferredPaymentMode, COUNT(\*) AS ActiveCustomerCount

FROM customer\_churn

WHERE ChurnStatus = 'Active'

GROUP BY PreferredPaymentMode

ORDER BY ActiveCustomerCount DESC

LIMIT 5;

-- 8. List the preferred login device(s) among customers who took more than 10 days

-- since their last order.

SELECT DISTINCT PreferredLoginDevice

FROM customer\_churn

WHERE DaySinceLastOrder > 10;

-- 9. List the number of active customers who spent more than 3 hours on the app.

SELECT COUNT(\*) AS ActiveCustomerCount

FROM customer\_churn

WHERE ChurnStatus = 'Active'

AND HoursSpentOnApp > 3;

-- 10. Find the average cashback amount received by customers who spent at least 2

-- hours on the app.

SELECT AVG(CashbackAmount) AS AverageCashback

FROM customer\_churn

WHERE HoursSpentOnApp >= 2;

-- 11. Display the maximum hours spent on the app by customers in each preferred

-- order category.

SELECT PreferedOrderCat, MAX(HoursSpentOnApp) AS MaxHoursSpent

FROM customer\_churn

GROUP BY PreferedOrderCat;

-- 12. Find the average order amount hike from last year for customers in each marital

-- status category.

SELECT MaritalStatus, AVG(OrderAmountHikeFromLastYear) AS AverageOrderAmountHike

FROM customer\_churn

GROUP BY MaritalStatus;

-- 13. Calculate the total order amount hike from last year for customers who are single

-- and prefer mobile phones for ordering.

SELECT SUM(OrderAmountHikeFromLastYear) AS TotalOrderAmountHike

FROM customer\_churn

WHERE MaritalStatus = 'Single'

AND PreferedOrderCat = 'Mobile Phone';

-- 14. Find the average number of devices registered among customers who used UPI as

-- their preferred payment mode.

SELECT AVG(NumberOfDeviceRegistered) AS AveragNumberOfDeviceRegistered

FROM customer\_churn

WHERE PreferredPaymentMode = 'UPI';

select \* from customer\_churn;

-- 15. Determine the city tier with the highest number of customers.

SELECT CityTier, COUNT(\*) AS CustomerCount

FROM customer\_churn

GROUP BY CityTier

ORDER BY CustomerCount DESC

LIMIT 3;

-- 16. Find the marital status of customers with the highest number of addresses.

SELECT MaritalStatus, COUNT(NumberOfAddress) AS NumberOfAddressCount

FROM customer\_churn

GROUP BY MaritalStatus

ORDER BY NumberOfAddressCount DESC

LIMIT 1;

select \* from customer\_churn;

-- 17. Identify the gender that utilized the highest number of coupons.

SELECT Gender, COUNT(CouponUsed) AS CouponCount

FROM customer\_churn

WHERE CouponUsed IS NOT NULL

GROUP BY Gender

ORDER BY CouponCount DESC

LIMIT 2;

-- 18. List the average satisfaction score in each of the preferred order categories.

SELECT PreferedOrderCat, AVG(SatisfactionScore) AS AverageSatisfactionScore

FROM customer\_churn

GROUP BY PreferedOrderCat;

-- 19. Calculate the total order count for customers who prefer using credit cards and

-- have the maximum satisfaction score.

SELECT COUNT(\*) AS TotalOrderCount

FROM customer\_churn

WHERE PreferredPaymentMode = 'Credit Card'

AND SatisfactionScore = (SELECT MAX(SatisfactionScore) FROM customer\_churn);

-- 20. How many customers are there who spent only one hour on the app and days

-- since their last order was more than 5?

SELECT COUNT(\*) AS CustomerCount

FROM customer\_churn

WHERE HoursSpentOnApp = 1

AND DaySinceLastOrder > 5;

-- 21. What is the average satisfaction score of customers who have complained?

SELECT AVG(SatisfactionScore) AS AverageSatisfactionScore

FROM customer\_churn

WHERE ComplaintReceived = 'Yes';

-- 22. How many customers are there in each preferred order category?

SELECT PreferedOrderCat, COUNT(\*) AS CustomerCount

FROM customer\_churn

GROUP BY PreferedOrderCat;

-- 23. What is the average cashback amount received by married customers?

SELECT AVG(CashbackAmount) AS AverageCashback

FROM customer\_churn

WHERE MaritalStatus = 'Married';

-- 24. What is the average number of devices registered by customers who are not using

-- Mobile Phone as their preferred login device?

SELECT AVG(NumberOfDeviceRegistered) AS AverageNumberOfDeviceRegistered

FROM customer\_churn

WHERE PreferredLoginDevice <> 'Mobile Phone';

select \* from customer\_churn;

-- 25. List the preferred order category among customers who used more than 5 coupons.

SELECT PreferedOrderCat, COUNT(\*) AS CouponCount

FROM customer\_churn

WHERE CouponUsed > 5

GROUP BY PreferedOrderCat;

-- 26. List the top 3 preferred order categories with the highest average cashback amount.

SELECT PreferedOrderCat, AVG(CashbackAmount) AS AverageCashback

FROM customer\_churn

GROUP BY PreferedOrderCat

ORDER BY AverageCashback DESC

LIMIT 6;

-- 27. Find the preferred payment modes of customers whose average

-- tenure is 10 months and have placed more than 500 orders.

SELECT PreferredPaymentMode

FROM customer\_churn

GROUP BY PreferredPaymentMode

HAVING AVG(OrderAmountHikeFromlastYear) = 10

AND COUNT(CustomerID) > 500;

-- 28. Categorize customers based on their distance from the warehouse to home.

-- Then, display the churn status breakdown for each distance category.

SELECT CASE

WHEN WarehouseToHome <= 5 THEN 'Very Close Distance'

WHEN WarehouseToHome <= 10 THEN 'Close Distance'

WHEN WarehouseToHome <= 15 THEN 'Moderate Distance'

ELSE 'Far Distance'

END AS DistanceCategory, ChurnStatus, COUNT(\*) AS CustomerCount

FROM customer\_churn

GROUP BY DistanceCategory, ChurnStatus;

-- 29. List the customer’s order details who are married, live in City Tier-1,

-- and their order counts are more than the average number of orders placed by all customers.

WITH AverageOrderCount AS (

SELECT AVG(OrderCount) AS AvgOrderCount

FROM customer\_churn

)

SELECT \* FROM customer\_churn

WHERE MaritalStatus = 'Married'

AND CityTier = 'Tier-1'

AND OrderCount > (SELECT AvgOrderCount FROM AverageOrderCount);

-- 30.a: Create the customer\_returns Table and Insert Data

-- Create the 'customer\_returns' table

CREATE TABLE ecomm.customer\_returns (

ReturnID INT PRIMARY KEY,

CustomerID INT,

ReturnDate DATE,

RefundAmount DECIMAL(10, 2)

);

-- Insert the data into 'customer\_returns'

INSERT INTO ecomm.customer\_returns (ReturnID, CustomerID, ReturnDate, RefundAmount) VALUES

(1001, 50022, '2023-01-01', 2130),

(1002, 50316, '2023-01-23', 2000),

(1003, 51099, '2023-02-14', 2290),

(1004, 52321, '2023-03-08', 2510),

(1005, 52928, '2023-03-20', 3000),

(1006, 53749, '2023-04-17', 1740),

(1007, 54206, '2023-04-21', 3250),

(1008, 54838, '2023-04-30', 1990);

-- 30.b: Display Return Details with Customer Details for Churned Customers Who Have Made Complaints

SELECT cr.ReturnID,cr.CustomerID,cr.ReturnDate,cr.RefundAmount,cc.\*

FROM ecomm.customer\_returns cr

JOIN customer\_churn cc ON cr.CustomerID = cc.CustomerID

WHERE cc.ChurnStatus = 'Churned'

AND cc.ComplaintReceived = 'Yes';

**Conclusion**

The analysis of customer churn in the e-commerce sector is essential for businesses aiming to improve retention rates and enhance overall customer satisfaction. Throughout this project, we have explored a rich dataset containing various attributes that influence customer behavior and decision-making. By systematically processing and analyzing this data, we were able to uncover valuable insights into the factors that contribute to customer churn.

**Key Findings**

1. \*\*Churn vs. Active Customers\*\*: Our exploration highlighted the distribution of churned and active customers, providing a clearer understanding of the customer base and the scope of potential churn issues that need addressing.

2. \*\*Tenure and Satisfaction\*\*: We found a strong correlation between customer tenure and churn, indicating that longer-tenured customers tend to exhibit higher loyalty. Additionally, satisfaction scores significantly influenced churn behavior, reinforcing the need for businesses to prioritize customer satisfaction initiatives.

3. \*\*Impact of Complaints\*\*: The analysis revealed that customers who complained were more likely to churn. This finding underscores the importance of not only addressing complaints swiftly but also proactively improving the overall customer experience to reduce the likelihood of churn.

4. \*\*Spending Patterns\*\*: By investigating spending behaviors, we determined spending patterns related to churn, including the total cashback earned by churned customers. This insight can guide marketing promotions and retention strategies aimed at enhancing customer engagement.

5. \*\*Demographic Insights\*\*: Gender and marital status analyses provided nuanced understanding of customer preferences and behaviors, which can inform targeted marketing campaigns.

6. \*\*Payment Preferences and Order Categories\*\*: The preferred payment modes and order categories exemplified variations between active and churned customers, suggesting that aligning offerings with customer preferences can help reduce attrition.

7. \*\*Geographic Insights\*\*: Identifying the city tiers or regions with the highest churn rates can help businesses focus their retention efforts where they are most needed.

**Recommendations**

Based on the findings from this analysis, several recommendations can be made to e-commerce businesses:

- \*\*Enhance Customer Support\*\*: Implement a robust customer service framework to address complaints promptly and effectively, which can help mitigate churn.

- \*\*Improve Customer Engagement\*\*: Use targeted marketing strategies that consider customer demographics and preferences to enhance loyalty and reduce the likelihood of churn.

- \*\*Leverage Data Analytics\*\*: Continuous monitoring of customer behavior, preferences, and feedback should be integrated into business practices to identify at-risk customers swiftly and take preemptive action.

- \*\*Promote Customer Retention Programs\*\*: Establish loyalty programs or perks that reward long-term customers, cultivating deeper relationships with them.

- \*\*Tailor Marketing Efforts\*\*: Utilize insights from satisfaction scores and spending habits to create personalized marketing campaigns that resonate with specific customer segments.

In conclusion, the insights derived from this e-commerce customer churn analysis empower businesses to make informed decisions that enhance customer retention and satisfaction. By addressing the factors contributing to churn and implementing targeted strategies, e-commerce companies can foster loyalty, drive revenue growth, and secure a competitive edge in an ever-evolving market landscape. The findings from this project will serve as a valuable resource for stakeholders and decision-makers seeking to optimize customer retention strategies while maintaining a focus on customer-centric practices.